DOCUMENT RESUME

ED 098 539 CS 001 479

AUTHOR Walter, Richard B.
TITLE History and Development of the Informal Reading Inventory.
PUB DATE 74
NOTE 18p.; Unpublished study prepared at Kean College of New Jersey

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Educational History; Grouping (Instructional Purposes); *Informal Reading Inventory; *Literature Reviews; *Reading Diagnosis; Reading Instruction; *Reading Tests; Teacher Education

ABSTRACT

An informal reading inventory (IRI) consists of graded passages of increasing difficulty and a series of comprehension questions for each grade level reading selection. The IRI is most often used to assist classroom teachers in the placement of children in groups for purposes of instruction. In this paper, the history of the IRI is presented and the problem of selecting performance criteria, validity and reliability, and the value of IRIs for determining the instructional level of pupils are discussed. Selected literature, supporting the contention that most teachers cannot be successful in using the IRI without training in construction, administration, and interpretation of such an instrument, concludes the paper. (TO)

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Richard B. Walter
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

BEST COPY AVAILABLE

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ED 093539

History and Development of the Informal Reading Inventory

Richard B. Walter

It is appropriate to begin the discussion of IRIs with Betts' initial article on this subject. Betts emphasized the use and value of the IRI in assisting classroom teachers in the placement of children in groups for purposes of instruction.[1]

Betts stated the following advantages of using an IRI:

1. The teacher is given direct evidence on achievement and needs in terms of available instructional material.
2. The teacher is provided with a technique for detecting everyday needs in the classroom.
3. The child is convinced of his needs and sees how to improve his skills.[2]

Basically, an IRI consists of a series of graded passages, of increasing difficulty, and a series of comprehension questions for each grade level selection to be read. Betts has suggested the following criteria for the selection and organization process of the passages to be included in an IRI.

> First, the material should be within the interests of the learner. When it is not possible to live up to this criterion, extra motivation must be supplied in the tests situation. Second, the material should be graded in readability. Third, the content of the material should be the same as, or closely related to, that available in the classroom situation.[3]

---

[1] Emmett A. Betts, "Reading Problems at the Intermediate Grade Level", Elementary School Journal, XL (June, 1940), 737-46.

[2] Emmett A. Betts, Foundations of Reading Instruction, (New York: American Book Company, 1954), p. 478.

[3] Ibid., pp. 457-58.

CS 001 426

BEST COPY AVAILABLE

Since Betts' initial work in the area of IRIs, the instrument has found widespread acceptance in the field of reading. Several other investigators have attempted to determine criteria for interpreting IRIs. The most significant studies are presented.

Killgallon's study of 1942 is highly important. His contribution was two-fold: first, he reported that standardized tests did not discriminate at lower levels of performance,[4] and, second, he furnished Betts with part of the criteria which Betts later adopted for the scoring of his IRI. Killgallon examined 41 fourth graders with an informal reading instrument and attempted to establish performance criteria for scoring such instruments. The procedure used was unorthodox. Criteria were prepared, the subjects tested, the performance analyzed, and a new set of criteria devised. This procedure yielded the 95 per cent word recognition score. It was this criterion score which was adopted by Betts.

Cooper also attempted to establish criteria for an IRI. Based on a pre-test/post-test design, Cooper

---

[4]Patsy A. Killgallon, "A Study to Determine Relationships Among Certain Pupils' Adjustments in Language Situations", (unpublished doctoral dissertation, Pennsylvania State College, 1942).

compared standardized test scores administered at the beginning and end of the experimental period and observed that the group which made the fewest errors in word perception made the greatest gains in overall reading achievement. The converse was also true.[5] However, Cooper's sample lacked the size and diversity to allow his results to be generalized to other populations without a degree of caution.

Cooper's study lends support to the assumption that a pupil's reading ability can be judged, in part, by the degree of word perception he has. Thus, IRIs typically assess the amount of word recognition errors at each grade level as an indication of performance.

McCracken's study appears to be the most thoroughly conducted investigation on this subject to date. The passages of the test were validated with accepted readability formulae, and they closely adhered to the vocabulary contents of three basal reader series. McCracken's sample was large, over 600, and encompassed grades one through six.[6] Although

---

[5]J. Louis Cooper, "The Effect of Adjustment of Basal Reading Achievement" (unpublished doctoral dissertation, Boston University, 1952).

[6]Robert A. McCracken, "The Development and Validation of the Standard Reading Inventory for the Individual Appraisal of Reading Performance in Grades One through Six, "Proceedings of the International Reading Association, IX (Newark, Delaware: International Reading Association, 1964), 310-13.

McCracken involved the ratings of twenty-five nationally known reading authorities, he did use Betts' criteria, partially based on Killgallon's study. Because of this, he has had his findings accepted with a degree of caution. The major point of contention has always been aimed at the acceptance of these performance criteria, and the subject has not yet been laid to rest.

Performance Criteria of the Informal Reading Inventory

As previously stated, the issue of performance criteria has always been unsettled. Interestingly enough, no one is actually aware of how Betts' and Killgallon's criteria were determined. Beldin asked Betts how they were determined and Betts replied by personal correspondence:

> This study (Killgallon's) was done after we had been using mimeographed materials, etc., for teaching the informal inventory. Hence, I would assume that he (Killgallon) had used criteria we had originally set up.[7]

---

[7] H.O. Beldin, "Informal Reading Testing: Historical Review and Review of the Research", Reading Difficulties: Diagnosis, Correction and Remediation (Newark, Delaware: International Reading Association, 1970), p. 75.

Killgallon's study contains conflicting evidence. Killgallon wrote:

> Criteria for defining the probable instructional level, the probable frustration level, and the probable reading capacity level were arbitrarily established by the investigator in connection with the Informal Reading Inventory.[8]

Although the criteria used by Killgallon and later adopted by Betts appear to be arbitrary decisions, at best, these criteria are not at all inconsistent with previously done research by Thorndike,[9, 10] Bolenius,[11] and Durrell.[12] Since all theories must have their beginnings, Killgallon is to be commended for having taken such a step. It should not be forgotten that these criteria have been fairly well validated through use by many of the more prominant authorities in the field of reading.

---

[8]Killgallon, op. cit., p. 9.

[9]E. L. Thorndike, "Improving the Ability to Read", Teachers College Record, XXXVI (November, 1934), 123-44.

[10]E. L. Thorndike, "Improving the Ability to Read", Teachers College Record, XXXVI (December, 1934), 229-41.

[11]E. M. Bolenius, Teachers Manual of Silent and Oral Reading (Boston: Houghton-Mifflin Co., 1919).

[12]Donald Durrell, "Individual Differences and Their Implications with Respect to Instruction in Reading", Teaching of Reading: A Second Report, The Thirty-sixth Yearbook of the National Society for the Study of Education, (Bloomington, Illinois: Public School Publishing Company, 1937), pp. 325-26.

Perhaps the most outspoken critic of the Betts-Killgallon performance levels has been Powell, who contended that the Betts-Killgallon criteria were unrealistically high and lacked the support of empirical data.[13] Powell compiled the performance criteria of various authorities and found little deviation from the Betts-Killgallon criteria. Notable exceptions to the statement came from Smith[14], who proposed a lower percentage of accuracy, and Spache, who asserted that Betts' standards were arbitrarily set too high.[15] Powell completed a study in which he tested the hypothesis that a suitable criterion for word recognition could be lower than the previously accepted 95 per cent level. He was concerned with two components: word recognition and comprehension. Holding comprehension constant at 70 per cent, Powell scanned the performance of his subjects on word recognition. He submitted the

---

[13]William R. Powell, "Reappraising the Criteria for Interpreting Informal Reading Inventories", Proceedings of the International Reading Association, XIII, Part IV (Newark, Delaware: International Reading Association, 1969), 100-09.

[14]Nila Banton Smith, Graded Selections for Informal Reading: Diagnosis for Grades 1 through 3, (New York: University Press, 1959).

[15]George D. Spache, Reading in the Elementary School (Boston: Allyn and Bacon, 1954).

following results:

> The data suggest that the mean scores of grades one and two tend to cluster together percentage-wise, and the mean percentages of grades three, four, five, and six, form a relatively similar percentage score. The data clearly indicate that pupils in grades one and two could tolerate on the average an 85 per cent word recognition score and still maintain 75 per cent comprehension. To say I found this astonishing is an understatement: The findings that pupils in grades three through six could tolerate on the average a 91 to 94 per cent word recognition score while maintaining 70 per cent comprehension is commensurate with the data of Killgallon.[16]

Many authors still continue to use various modifications of the traditional performance criteria. Ransom summarized the performance criteria of several authors to illustrate the diversity of opinion on that subject. That table is presented on the next page.

The preceding discussion of performance criteria and the chart are presented to illustrate that the IRI is a powerful concept which is accepted by reading authorities even though a difference of opinion exists as to which performance criteria should be accepted.

---

[16]Powell, op. cit., p. 106.

TABLE 1

CRITERIA FOR ESTABLISHING READING LEVELS WHEN USING AN INFORMAL READING INVENTORY[17]

Author	Independent		Instructional		Frustration	
	Word Recognition	Comprehension	Word Recognition	Comprehension	Word Recognition	Comprehension
Betts	99+	90+	95+	75+	90-	50-
Botel			95+	75+	95-	75-
Harris	97+		95+ to 98+		90-	
Wheeler and Smith	98+	75+	95+	75+	95-	75-
Williams	99+	90+	90+	70+		
Cooper			98+ 96+	70+ (Primary) 60+ (Intermediate)		

Numerical Values = Percentage Correct

[17] Ransom, op. cit., p. 19.

Validity and Reliability of the Informal Reading Inventory

The validity and reliability of an IRI is of an elusive nature; little has been written on this aspect of the IRI until recent years. For many years, the validity of the IRI has been assumed. Since the IRI is usually composed of selections from actual basal readers the pupil is using, his reading performance is a valid indication of his ability to handle basal material. Furthermore, since the questions associated with the graded passages are, in fact, similar to actual classroom questions, their validity is also assumed. Researxh in recent years has attempted to lend support to the assumptions of validity with empirical data. An example of such a study was done by McCracken, who statistically validated his Standardized Reading Inventory. The Inventory contains many features found in most IRIs. Using alternate forms of his instrument, he found significant correlations among them at the .001 level of confidence.[18] Sipay compared two parallel forms of an IRI and found correlations of .78 to .82 between them.[19]

---

[18]Robert A. McCracken, "The Development and Validation of the Standard Reafing Inventory for the Individual Appraisal of Reading Performance in Grades One through Six" (unpublished doctoral dissertation, Syracuse University, 1963), pp. 47-68.

[19]Edward R. Sipay, "A Comparison of Standardized Reading Scores and Functional Reading Levels", The Reading Teacher XVII (January, 1964), 265-68.

The question of the reliability of the IRI is, at best, a complex problem. The reliability technique of test-retest does not lend itself to an IRI. There are two reasons for this. The first reason stems from the probability that performance will improve with each additional exposure to the material. The second reason rests with the purpose of the instrument. An IRI is administered with the intent of being able to improve performance with a prescriptive reading program based upon the information obtained from the instrument. Thus, the instrument would rarely be given in an effort to record a duplicate level of performance. The technique of split-half reliability is not suitable for use since the values of the test items are unequal. It would seem, then, that the reliability of an IRI is best measured by comparing performance on this instrument with actual reading performance. Hence, the argument that the IRI is valid, since the materials used in the test construction are, ideally, the identical material used in actual classroom reading situations, that is, basal readers. It is further argued that because of this factor, IRIs are more accurate than standardized tests for purposes of determining a pupil's instructional level. The following review of the literature would seem to indicate that there is a good deal of truth to this argument.

Comparison of the Informal Reading Inventory with Standardized Tests as Determiners of Instructional Level

McCracken compared pupils' scores on an IRI with their scores on the Iowa Comprehension Test. He reported that:

> Four pupils achieved 0 to 1 year better on the Iowa Comprehension than their informal minimum instructional level; twelve pupils achieved 1 to 2 years better; twenty-eight pupils achieved 2 to 3 years better: nine pupils achieved 3 to 4 years better; and three pupils achieved 4 to 5 years better.[20]

On the vocabulary section, McCracken found that:

> Six children achieved informal word recognition levels higher than their Iowa vocabulary grade level, four by less than one year, and two by one year. Twenty-two pupils achieved 0 to 1 year better on the Iowa vocabulary than their informal vocabulary level; eighteen pupils achieved 1 to 2 years better; nine pupils 2 to 3 years better; and 1 pupil achieved 3 to 4 years better.[21]

He summarized:

> The standardized test scores would place 63% (35 pupils) of the children at a level of frustration if their scores were used to determine the book level needed for reading instruction. It would place 93% (52 pupils) of the children in a book which is too hard for pupils' and teacher comfort.[22]

McCracken concluded that "the average difference between the Iowa Reading Comprehension grade levels and the informal immediate instructional reading level was 2.3 years".[23]

---

[20] R. A. McCracken, "Standardized Tests and Informal Reading Inventories", Education, LXXXII (February, 1962), 367.

[21] Ibid.

[22] Ibid., p. 368.

[23] Ibid., p. 367.

Wheeler and Smith acknowledge the tendency for standardized tests to over-estimate a pupil's instructional level; they stated:

> It has long been known by those who work closely with children that grade placement scores on reading tests for primary grades often have little relationship to the child's actual instructional reading level. In the Reading Clinic (University of Miami) we have found that pupils who cannot read a primer will often score above the second grade reading level on some standardized tests.[24]

Betts stated that "standardized tests tend to place children at their frustration level".[25]

In a study similar to McCracken's, Sipay concluded that:

> All three standardized tests Metropolitan Reading Test (1960), Gates Reading Survey (1958), and the California Reading Test (1957), tended to overestimate the instructional level by approximately one or more grade levels.[26]

And Farr asserted, "The use of standardized reading tests as indicators of instructional reading level should be abandoned".[27]

This suggestion is no doubt based upon the fact that standardized tests are generally less accurate at the

---

[24]Lester R. Wheeler and Edwin H. Smith, "Modification of the Informal Reading Inventory", Elementary English, XXXIV (April, 1957), 224.

[25]Emmett A. Betts. Foundations of Reading Instruction (New York: American Book Company, 1957), pp. 450-51.

[26]Sipay, op. cit., p. 267

[27]Roger Farr, "Reading Tests and Teachers", Educational Resources Information Center Document ED 020 082 (Bethesda, Maryland: ERIC Document Reproduction Service, 1968), Abstract.

extremes of performance. In a study in which Betts compared the grade equivalent scores of several standardized tests designed for use at the fifth grade level with actual reading performance, he concluded that standardized tests are less than adequate for determining achievement levels for pupils at either end of the distribution of scores.[28] Chall also pointed out that standardized tests designed for a few grade levels gave a poor indication of reading performance for the poorest and the most able readers.[29]

The above is by no means the total of all the available research which supports the use of the IRI rather than the standardized test for purposes of determining instructional level, but the implication is clear: standardized tests are not accurate for this purpose. This being the case, it seems that the obvious conclusion is for classroom teachers to administer an IRI to each pupil on an individual basis. This, however, would not be practical for a number of reasons, the most obvious of which is the time factor. The testing of an entire class would extend

---

[28]Betts, Foundations of Reading Instruction, op. cit., p. 441.

[29]Jeanne S. Chall, "Interpretation of Results of Standardized Reading Tests", Evaluation of Reading, Supplementary Educational Monographs, No. 88 (Chicago: University of Chicago Press, 1958), 133-38.

well into the beginning of the school year, and the instructional levels of many pupils would not be ascertained in time to be helpful to the teacher. This is not, however, the only disadvantage. The following discussion will introduce opinion as to some additional problems inherent in the use of the IRI.

Problems Inherent in the Use of the Informal Reading Inventory

There is reason to believe that the average untrained classroom teacher does not have the skills to construct, administer, and interpret an IRI. Even if allowances are made for the use of commercially available IRIs which have been carefully constructed, the tasks of administration and interpretation still remain beyond the skills of untrained teachers.

Kender addressed himself to this point and maintained that:

> Anyone who is expected to administer an informal reading inventory must be thoroughly knowledgeable about the reading process and thoroughly skilled in administering the instrument.[30]

---

[30]Joseph P. Kender, "How Useful Are Informal Reading Tests?", Journal of Reading, XI (February, 1968), 341.

Levy supported Kender's contention in her reply to his article. She stated that "the informal tests take a great deal of time as well as knowledge".[31]

Kender specifically stated that:

> The usefulness, then, of an informal reading test is in direct proportion to the knowledge of the examiner who uses it; therefore, it is unlikely that just any classroom teacher can easily administer an informal reading inventory and judiciously interpret its results as is sometimes claimed.[32]

Kelley's thoughts on this subject were that:

> Even though the informal reading inventory represents a most effective instrument for evaluating a reader's performance, perhaps its greatest disadvantage relates to the competency of the examiner, since the accuracy of the data received through its use depends almost entirely upon the competence of the examiner.[33]

As a result of her study, Ladd decided that teachers were often inaccurate in identifying oral reading errors. Ladd's conclusions lend support to previously stated opinions which suggested that a high degree of skill is needed in the administration and interpretation of an

---

[31]B. K. Levy, "How Useful Are Informal Reading Tests?", Journal of Reading, XII (October, 1968), 38.

[32]Kender, op. cit., p. 341.

[33]Kelley, loc. cit.

informal reading inventory. She found that even after thirty hours of training, teachers in her study still incorrectly identified from 33 to 37 per cent of oral reading errors made by pupils.[34]

A further comment on this problem was expressed by Della-Piana, Jenson, and Murdock who concluded:

> It is our judgment that construction, administration, and interpretation of informal reading inventories are far too time consuming for the regular classroom teacher. She either will not learn the skills, or having learned them will not use them in practice because of competing demands upon her time.[35]

Thus, the problem of providing classroom teachers with a simply constructed, easily administered, and objectively scored instrument for determining instructional level has not been solved.

It seems that while an IRI yields an indication of instructional level, its major usefulness is as a diagnostic instrument. By carefully noting, classifying, and interpretating oral reading and comprehension errors, the

---

[34]Eleanor Mary Ladd, "A Comparison of Two Types of Training with Preference to Developing Skill in Diagnostic Oral Reading Testing", (unpublished doctoral dissertation, Florida State University, 1961).

[35]Gabriel Della-Piana, Betty Jenson, and Everett Murdock, "New Directions for Informal Reading Assessment", Reading Difficulties: Diagnosis, Correction, and Remediation (Newark, Delaware: International Reading Association, 1970), p. 127.

examiner is able to determine specifically the inadequate skills which result in poor reading performance. These observations should then result in a prescriptive program to improve reading performance. In this capacity, the IRI should not be discounted nor should it be replaced for this purpose. This clarification is offered since the major emphasis has been upon the IRI as an indicator of instructional level. Powell stated this most appropriately:

> The value of an IRI lies not in its identification of what has been called the instructional level (and the other levels by interpolation) because there are probably more effective and efficient methods of accomplishing such tasks. The use of cloze procedure is one alternative already available that has a considerable body of research data to support it. The real value of the IRI is that it affords the possibility of evaluating reading behavior in depth.[36]

Powell's statement is especially significant due to his reference to the cloze procedure as a possible substitute technique for determining the instructional levels of pupils.

Summary of the History & Development of the Informal Reading Inventory

The material included in the preceding section pertained to the history and development of the IRI. The

---

[36]William R. Powell, "The Validity of the Instructional Reading Level: Informal Reading Inventory", Proceedings of the International Reading Association, XV (Newark, Delaware: International Reading Association, 1971), 131.

problem of selecting performance criteria, validity and reliability, and the value of IRIs for determining the instructional level of pupils is also discussed. The final portion includes selected literature to support the contention that the average classroom teacher probably could not be successful in using the IRI due to a lack of training necessary to enable her to construct, administer, and interpret such an instrument.